VOLUME 6
NUMBER 4
DECEMBER 1990

ISSN 1016-2364

Journal of Information Science and Engineering

Special Issue on Computer Vision

INSTITUTE OF INFORMATION SCIENCE
ACADEMIA SINICA
Taipei, Taiwan, Republic of China

JOURNAL OF INFORMATION SCIENCE AND ENGINERRING 6, 325-337 (1990)

Environmental Studies of ICM Segmentation Algorithm

CHAUR-CHIN CHEN
Department of Computer Science
National Tsing Hua University
Hsinchu 30043, Taiwan, R.O.C.

RICHARD C. DUBES
Computer Science Department
Michigan State University
East Lansing, Michigan 48824, U.S.A.

The ICM (Iterated Conditional Modes) algorithm has recently been exploited for image segmentation. The ICM segmentation algorithm iteratively updates the label of each pixel until a prescribed criterion is achieved. Most experiments and work have assumed that the true labeling is modeled by a discrete Markov random field and that the observed degraded image is formed by adding i.i.d. Gaussian noise to the true image. This paper reports on the ICM algorithm with various assumptions of degradation models. We characterize the mathematical formulas, list the ICM algorithm, and give the experiments based on known model parameters to segment synthetic images. The ICM algorithm segments images reasonably well under a variety of degradation models even though prior information is inadequate. A practical application of ICM algorithm for reconstructing an infrared image for target recognition is given.

Keywords: Gibbs distribution, ICM, Markov random field, pixel labeling, segmentation.

1. INTRODUCTION

Let an M×N intensity image be defined as a coloring on an M×N lattice with possible colors 0, 1, ..., 255. Given an M×N intensity image, the problem is to label each site or pixel in this M×N lattice to optimize some prescribed criterion. The problem is either called image restoration [2, 9] or image segmentation [6, 7]. In image restoration, the labels in the true image can be chosen from {0, 1, ..., 255}. In image segmentation, the true labels can be chosen from A = {0, 1, ..., G-1}, where $G << 255$. For example, the labels can refer to a land-use category. In edge detection, there are only two possible labels (edge or non-edge). Thus, the edge detection problem can be viewed as a special case of image segmentation with $G = 2$.

Received December 15, 1989; revised July 15, 1990.
Communicated by Jun S. Huang.
Dr. Chen was partially supported by NSC Grant 79-0408-E007-21.
Dr. Dubes was supported by NSF Grant IRI-8901513.
Part of this paper was presented in Taiwan-CVGIP workshop in 1989.

This paper deals with image segmentation. A segmentation algorithm requires information about the image. A segmentation algorithm good for one image may not be appropriate for other images. A good review of segmentation techniques up to 1981 can be found in [15]. Haralick and Shapiro [10] also surveyed segmentation techniques with emphasis on edge detection.

Image segmentation using a Markov random field (MRF) to capture contextual information has been widely reported [2, 5-7, 9, 12-14]. Most papers either concentrate on the results of segmenting individual images or emphasize statistical analysis of image data. MRF model-based segmentation algorithms have had at least one of the following drawbacks making duplication of results difficult: (i) the mathematical formulas were not given exactly, (ii) the algorithms were not listed, (iii) the parameters used in the algorithms were not clearly specified, (iv) the image data were not sufficiently described.

Dubes et al. [8] have recently demonstrated that the ICM algorithm is a powerful MRF-Based labeling algorithm for image segmentation. In all of their experiments, the algorithm itself assumed that the degraded images were formed by adding i.i.d. Gaussian noise to each pixel of the true images, although some images involve correlated noise and textured regions. This paper reviews and extends the ICM algorithm to various models of degradation. We not only emphasize the mathematical results but also report experimental results and demonstrate its application to an infrared image.

2. BACKGROUND

The ICM algorithm attempts to label a lattice by locally optimizing the a posteriori distribution [2, 8]. An MRF is imposed as the prior distribution and a degradation model is assumed so that the a posteriori distribution is also an MRF. Fig. 1(a) defines a notation for the relative neighbors of pixel t while Fig. 1(b) shows the orders of neighbors of pixel t up to order 5 [4].

t:-11	t:-7	t:-6	t:+8	t:+12
t:-9	t:-3	t:-2	t:+4	t:+10
t:-5	t:-1	t	t:+1	t:+5
t:-10	t:-4	t:+2	t:+3	t:+9
t:-12	t:-8	t:+6	t:+7	t:+11

(a)

5	4	3	4	5
4	2	1	2	4
3	1	t	1	3
4	2	1	2	4
5	4	3	4	5

(b)

Fig. 1. The pixels and orders of neighbors of pixel t.

In the following discussion, we assume that the true labeling is a realization of an MRF whose Gibbs distribution is given by

$$f(\mathbf{x}) = e^{-U(\mathbf{x})}/Z,$$

where $\mathbf{x}$ is an MN-tuple vector, Z is the normalizing constant, and the energy function U is defined as

$$U(\mathbf{x}) = \sum_{t=1}^{MN} \sum_{r=1}^{c} \beta_r J(x_t, x_{t:+r}), \tag{1}$$

where $J(a,b) = -1$ if $a=b$, 0 if $a \neq b$; $c = 2$ for a 1st-order MRF and $c = 4$ for a 2nd-order MRF. For details refer to [1, 4, 11].

The observed images are obtained by degrading the MRF with the schemes described below.

Case 1

Each pixel is degraded with i.i.d. Gaussian noise. In other words, the observed image, denoted by a random vector $\mathbf{Y}$, is assumed to come from a known distribution conditioned on $\mathbf{X} = \mathbf{x}$, whose distribution is given by

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{MN} f(y_i|x_i). \tag{2}$$

Under the assumptions of (1) and (2), the a posteriori distribution of $\mathbf{X}$ conditioned on $\mathbf{y}$ $(\mathbf{X}|\mathbf{y})$ also defines an MRF [Appendix A]. In the case where

$$f(y_i|x_i) \sim N(\mu_{x_i}, \sigma_{x_i}^2), \tag{3}$$

the distribution of $\mathbf{X}|\mathbf{y}$ and its energy function are characterized by

$$f(\mathbf{x}|\mathbf{y}) = e^{-U(\mathbf{x}|\mathbf{y})}/Z_{\mathbf{x}|\mathbf{y}}, \text{ where } Z_{\mathbf{x}|\mathbf{y}} \text{ is the normalizing constant, and}$$

$$U(\mathbf{x}|\mathbf{y}) = \sum_{t=1}^{MN} \left[\frac{1}{2} \ln(\sigma_{x_t}^2) + \frac{(y_t - \mu_{x_t})^2}{2\sigma_{x_t}^2} + \sum_{r=1}^{c} \beta_r J(x_t, x_{t:+r}). \right] \tag{4}$$

Let $\mathbf{x}_{\partial t}$ be the collection of all neighbors of pixel t excluding pixel t. For example, in a 2nd-order neighborhood, $\mathbf{x}_{\partial t} = \{x_{t:\pm 1}, x_{t:\pm 2}, x_{t:\pm 3}, x_{t:\pm 4}\}$. A simple derivation according to Eq. (1) of Appendix A shows that

$$f(x_t|\mathbf{x}_{\partial t}, \mathbf{y}) = e^{-U_t}/Z_t, \text{ where } Z_t \text{ is a normalizing constant, and}$$

$$U_t = \frac{1}{2} \ln(\sigma_{x_t}^2) + \frac{(y_t - \mu_{x_t})^2}{2\sigma_{x_t}^2} + \sum_{r=1}^{c} \beta_r [J(x_t, x_{t:+r}) + J(x_t, x_{t:-r})]. \tag{5}$$

Case 2

Pixels are degraded by correlated noise generated from an auto-normal MRF model [1, 4] whose density function, conditioned on the true labeling, is given by

$$f(\mathbf{y}|\mathbf{x}) = \frac{|B|^{1/2}}{(2\pi\sigma^2)^{MN/2}} \exp[-(\mathbf{y} - \mu_x)^T B(\mathbf{y} - \mu_x)/2\sigma^2]. \tag{6}$$

The matrix B is an MN×MN block circulant matrix with M^2 blocks of N×N circulant matrices B_{ij}'s [3, 4] defiined below.

$$B = \begin{bmatrix} B_{11} & B_{12} & . & . & B_{1M} \\ B_{1M} & B_{11} & . & . & B_{1,M-1} \\ . & . & . & . & . \\ . & . & . & . & . \\ B_{12} & B_{13} & . & . & B_{11} \end{bmatrix}$$

For the 2nd-order neighborhood,

B_{11} = circulant $(1, -\theta_1, 0, 0, \ldots, 0, -\theta_1)$,

B_{12} = circulant $(-\theta_2, -\theta_3, 0, 0, \ldots, 0, -\theta_4)$,

B_{1N} = circulant $(-\theta_2, -\theta_4, 0, 0, \ldots, 0, -\theta_3)$,

$B_{1j} = O$ for $2 < j < N$.

The distribution $\mathbf{Y}|\mathbf{x}$ specified in Eq. (6) is also called a Gaussian Markov random field (GMRF). It can be shown that $f(\mathbf{x}|\mathbf{y})$ also defines an MRF whose energy function is given by

$$U(\mathbf{x}|\mathbf{y}) = \sum_{t=1}^{MN} \left[\frac{1}{2\sigma^2} \{(y_t - \mu_{x_t})^2 - 2 \sum_{r=1}^{c'} \theta_r (y_t - \mu_{x_t})(y_{t:+r} - \mu_{x_{t:+r}})\} + \sum_{r=1}^{c} \beta_r J(x_t, x_{t:+r}) \right] \tag{7}$$

Note that the size of neighborhoods for $\mathbf{Y}|\mathbf{x}$ and $\mathbf{X}$, c' and c, need not be the same, but one must be dominated by the other to guarantee that $f(\mathbf{x}|\mathbf{y})$ defines an MRF. Without loss of generality, we assume that $c' = c$ in this paper. A further simplification shows that the conditional density, X_t, the label of pixel t, given its neighbors can be expressed as

$$f(x_t|\mathbf{x}_{\partial t}, \mathbf{y}) = e^{-U_t}/Z_t, \text{ where } Z_t \text{ is a normalizing constant, and}$$

$$U_t = \frac{1}{2\sigma^2} \{(y_t - \mu_{x_t})^2 - 2 \sum_{r=1}^{c} \theta_r (y_t - \mu_{x_t})[(y_{t:+r} - \mu_{x_{t:+r}}) + (y_{t:-r} - \mu_{x_{t:-r}})]\}$$

$$+ \sum_{r=1}^{c} \beta_r [J(x_t, x_{t:+r}) + J(x_t, x_{t:-r})]. \tag{8}$$

Case 3

The observed image **y** is an M×N multispectral image, so each pixel consists of a vector of the same length, say d. Our goal is still to find an M×N labeling **x** which maximizes the a posteriori distribution of **X** given **y**. We further assume that

$$f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{MN} f(\mathbf{y}_i|x_i), \text{ and } f(\mathbf{y}_i|x_i) \sim N(\mu_{x_i}, \mathbf{C}_{x_i}), \tag{9}$$

where $\mathbf{y}_\cdot$, $\mu_\cdot$ are d-tuple vectors and $\mathbf{C}_\cdot$ is a d×d positive definite matrix. Under the assumptions of (1) and (9), the a posteriori distribution of $\mathbf{X}|\mathbf{y}$ also defines an MRF whose distribution and energy function are characterized by

$$f(\mathbf{x}|\mathbf{y}) = e^{-U(\mathbf{x}|\mathbf{y})}/Z_{\mathbf{x}|\mathbf{y}}, \text{ where } Z_{\mathbf{x}|\mathbf{y}} \text{ is a normalizing constant, and}$$

$$U(\mathbf{x}|\mathbf{y}) = \sum_{t=1}^{MN}\left[\frac{1}{2}\{\ln|\mathbf{C}_{x_t}| + (\mathbf{y}_t - \mu_{x_t})^T\mathbf{C}_{x_t}^{-1}(\mathbf{y}_t - \mu_{x_t})\} + \sum_{r=1}^{c}\beta_r J(x_t, x_{t:+r})\right]. \tag{10}$$

A simplification shows that the conditional density can be expressed as

$$f(x_t|\mathbf{x}_{\partial t},\mathbf{y}) = e^{-U_t}/Z_t, \text{ where } Z_t \text{ is a normalizing constant, and}$$

$$U_t = \frac{1}{2}\{\ln|\mathbf{C}_{x_t}| + (\mathbf{y}_t - \mu_{x_t})^T\mathbf{C}_{x_t}^{-1}(\mathbf{y}_t - \mu_{x_t})\} + \sum_{r=1}^{c}\beta_r\{J(x_t, x_{t:+r}) + J(x_t, x_{t:-r})\}. \tag{11}$$

The ICM algorithm [2, 8] searches for an optimal labeling **x** based on an observed image **y** by iteratively minimizing the *conditional energy* given in Eqs. (5), (8), or (11). Notice that the ICM algorithm only guarantees finding a local optimal labeling. The environments of the ICM algorithm are described in the following Section.

3. PIXEL LABELING ALGORITHM-ICM

The ICM algorithm was first proposed by Besag in 1986 to find an optimal labeling **x** based on a given intensity image **y**. A prior MRF presents contextual information; for example, if all β_r's in Eq. (1) are positive and large enough, the MRF tends to generate homogeneous regions [4]. Given the initial labeling, it iteratively updates the label of pixel t by maximizing $f(x_t|\mathbf{x}_{\partial t},\mathbf{y})$ for t = 1, 2, ..., MN, where $\mathbf{x}_{\partial t} = \{x_{t:-r}, x_{t:+r}|r = 1, ..., c\}$. Note that the formula $f(x_t|\mathbf{x}_{\partial t},\mathbf{y}) = f(x_t|\mathbf{y},\tilde{\mathbf{x}}_t)$ was given in Eqs. (5), (8), or (11) according to different assumptions on $f(\mathbf{y}|\mathbf{x})$, where $\tilde{\mathbf{x}}_t = \{x_j|j = 1, 2, ..., MN, j \neq t\}$. The procedure is repeated until the local minimum energy is achieved. The initial labeling is usually obtained

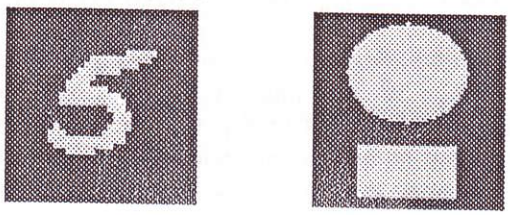
by applying a maximum likelihood estimation (MLE) individually on each pixel, namely, a classification without using spatial information. The ICM algorithm finds a labeling with a local minimum energy, it does not guarantee finding a labeling with a global minimum energy. Dubes et al. [8] demonstrated that under wrong prior information, labeling with a global minimum energy need not be the best solution. Besag [2] showed by experiments that it was usually enough to take 6 complete scans or fewer to achieve a reasonable labeling. The ICM algorithm is listed below and the experiments are given in the next Section.

ICM Algorithm

(a) Initialize a labeling by applying MLE for each pixel.
(b) For $t = 1$ to MN
$x_t \leftarrow g_0$ if $f(x_t = g_0 | \mathbf{x}_{\partial t}, \mathbf{y}) > f(x_t = g | \mathbf{x}_{\partial t}, \mathbf{y})$ for all $g \in A$ and $g_0 \in A$.
(c) Repeat (b) until "the energy achieves a local minimum" (6 scans here).
(d) $\mathbf{x}$ is the required labeling.

4. EXPERIMENTS

We now report on experiments showing how the ICM segmentation algorithm segments the degraded images obtained with the degradation processes given in Section 2. To let a boundary pixel have the same number of neighbors as an interior pixel, we assume that all images have periodic boundaries. Throughout the experiments, all of the images had size 64×64, and the isotropic 2nd-order Ising model [11] was used as the prior distribution, i.e., $c = 4$, and $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta$ in Eq. (1), where $\beta = 1.5$. The stopping criterion was defined to be 6 raster scans. In each case, our goal was to find the binary segmentation which had a local minimum of energy defined in Eqs. (4), (7), or (10). Two *true labelings*, each consisting of two labels 0 and 1, were considered throughout all of the experiments. The images displayed in Fig. 2 have gray levels 100 and 240 corresponding to labels 0 and 1, respectively, for display purposes. In the following experiments, we assume that the *true images* have gray levels 130 and 160 corresponding to labels 0 and 1 to avoid gray values beyond the display range [0, 255].

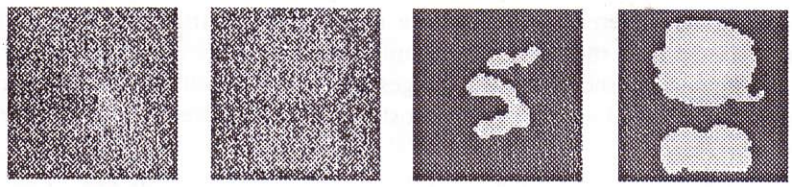


(a) (b)

Fig. 2. The Display of Images Corresponding to True Labelings.

Experiment-1 Degradation with i.i.d. Gaussian Noise

The degraded images shown in Figs. 3(a) and 3(b) are obtained by adding i.i.d. Gaussian noise with mean 0 and variance 900 to each pixel of the images in Fig. 2. Both of the gray level histograms of images in Fig. 3 are unimodal, so simple thresholding techniques are not appropriate [15]. We apply the ICM algorithm by using $f(x_t|\mathbf{x}_{\partial t}, \mathbf{Y})$ in Eq. (5) with known parameters to segment the degraded images in Fig. 3. The segmentation results shown in Figs. 3(c) and 3(d) are visually similar to the *true images* in Fig. 2, although not perfect.

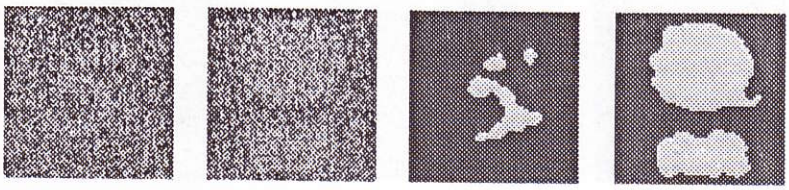


(a) (b) (c) (d)

Fig. 3. The Degraded and Segmented Images in *Experiment-1*.

Experiment-2 Degradation with Correlated Noise

The degraded images shown in Figs. 4(a) and 4(b) were obtained by adding the correlated noise generated from the Gaussian Markov random field [Appendix B, 1-4] with parameters $\mu = \mathbf{0}$, $\sigma = 30$, $\theta_1 = \theta_2 = 0.10$, $\theta_3 = \theta_4 = -0.05$ to the true images. Again, the gray level histograms of these images are unimodal, so simple thresholding techniques without using spatial information will lead to bad segmentation results. We applied the ICM algorithm by assuming $f(x_t|\mathbf{x}_{\partial t}, \mathbf{y})$ in Eq. (8) with known parameters to the degraded images in Figs. 4(a) and 4(b). The segmentation results shown in Figs. 4(c) and 4(d) are still reasonable since the signal to noise ratio is low (≈ 1). But the results in this Experiment are not as good as those in *Experiment-1*; we suspect that the wrong prior model (the true labeling is not a realization of the prior MRF model) associated with the correlated noise may attenuate the segmentation results.



(a) (b) (c) (d)

Fig. 4. The Degraded and Segmented Images in *Experiment-2*.

Experiment-3 Multispectral Images

This experiment examines the extension of ICM to multispectral images. We assume the true image is a 2-band *binary* multispectral image, and each pixel consists of a 2-tuple vector whose components have two possible values. The two possible vectors corresponding to each pixel are (100, 130) and (120, 160). The band-1 true image consists of gray levels 100 and 120. The band-2 true image consists of gray levels 130 and 160. The degraded images are obtained by adding i.i.d. multivariate normal noise with $\mu_0 = \mu_1 = 0$, $\mathbf{C}_0 = \mathbf{C}_1 = \begin{bmatrix} 400 & 200 \\ 200 & 900 \end{bmatrix}$ to each pixel of a true 2-band image. Figs. 5(a) and 5(b) show the degraded images corresponding to different bands whose corresponding true labeling is given in Fig. 2(a). We applied the ICM algorithm by assuming $f(x_t | \mathbf{x}_{\partial t}, y)$ in Eq. (11) with known parameters to the degraded images in Figs. 5(a) and 5(b). The segmentation results shown in Fig. 5(c) are almost perfect compared with the true image given in Fig. 2(a).

We repeat the process mentioned in this Experiment on the 2nd image whose *true labeling* corresponds to Fig. 2(b). The degraded images corresponding to two bands are given in Figures 5(d) and 5(e), respectively. The segmentation result shown in Fig. 5(f) are again almost perfect compared with the true image given in Fig. 2(b).

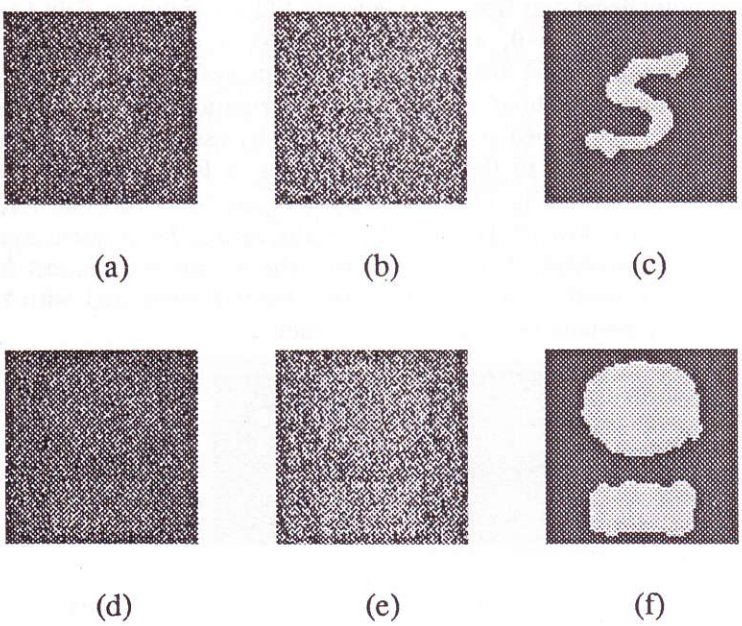


Fig. 5. The Degraded and Segmented Images in *Experiment-3*.

Experiment-4 Application to a Real Image

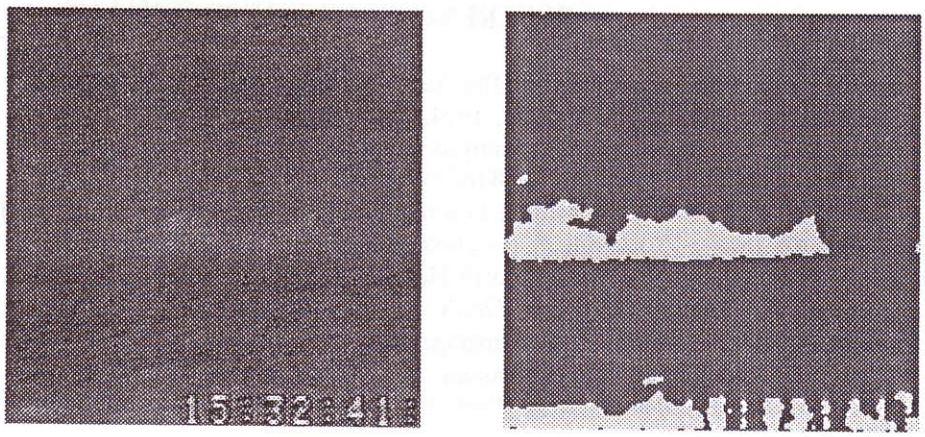
This experiment demonstrates the practicality of an ICM algorithm in the real world. Fig. 6(a) is a 200×200 infrared image with possible gray leves from 0 to 255. The histogram of the gray levels is unimodal. An ICM algorithm with estimated parameters used to segment and localize the object is described below.

ICM Algorithm with Estimated Parameters

(a) Let the histogram of a gray-level image take values from $[\mu_0, \mu_1]$, and let σ_0 and σ_1 be α-percent and γ-percent quantile of histogram, respectively.
(b) Find an initial $\mathbf{x}$ by applying MLE to label each pixel without using contextual information (i.e., $\beta=0$ in Eq. (5), (8), (11)). Let $\beta=1.5$.
(c) For $t=1$ to MN
$x_t \leftarrow g_0$ if $f(x_t=g_0|\mathbf{x}_{\partial t},\mathbf{y}) > f(x_t=g|\mathbf{x}_{\partial t},\mathbf{y})$ for all $g\in A$ and $g_0\in A$ using Eq. (5).
(d) Update means and variances μ_i, σ_i, $i=0, 1$, by MLE, based on $\mathbf{y}$ and current $\mathbf{x}$.
(e) Repeat (c) K iterations until "the energy achieves a local minimum."
(f) $\mathbf{x}$ is the output binary image.

The choice of (α, γ, K) depends on the signal to noise ratio. It is inappropriate to try to find a mathematical optimal (α, γ, K). In this experiment, we heuristically choose $\alpha=10$, $\gamma=90$, and $K=6$.

Note that this algorithm is developed for binary segmentation. For segmentation with more than two labels, step (a) may be modified by a clustering algorithm; see [14] for an instance.



(a) (b)

Fig. 6. The Infrared Image and the Result of the ICM Algorithm.

The result of applying the ICM algorithm with estimated parameters to the image in Fig. 6(a) is displayed in Fig. 6(b). The result is encouraging even if neither the prior nor the degradation model is suitable for the image.

5. DISCUSSION AND FUTURE RESEARCH

The existing MRF model-based segmentation algorithms [1, 7, 9, 12, 13] are restricted to using images which are degraded by adding independent noise to each pixel. We considered more flexible degradation models. We have derived the mathematical formulas for the ICM algorithm according to various assumptions of the degradation model and have demonstrated by experiments for both artificial and infrared images how the ICM algorithm can be used for image segmentation. The ICM algorithm can be easily adapted as other priors and degradation models $f(\mathbf{x})$ and $f(\mathbf{y}|\mathbf{x})$ according to the theorem given in Appendix A.

Although the experiments are restricted to segmentation for binary labels, the extension to multi-level labelings is straightforward but needs further study. It must be mentioned that the true labeling may never be a realization of the prior model. Moreover, it is not necessary to assume that the prior one is an isotropic Ising model [4, 11]; other discrete MRF models [6] may also be used. The choice of the best prior model is both image-dependent and goal-oriented.

We summarize some possible research in the future as follows.

(1) The extension of the ICM algorithm to segment images with more than two labels.
(2) The effect of prior and degradation models on synthetic and real images.
(3) The application of the ICM-based algorithm to edge detection and boundary extraction.
(4) Applications to real images such as X-ray images, ultrasonic images, and microphotographs.

REFERENCES

1. Besag, J., "Spatial Interaction and the Statistical Analysis of Lattice Systems," *J. Royal Stat. Soc. Ser.B,* Vol. 36, 1974, pp. 192-236.
2. Besag, J., "On the Statistical Analysis of Dirty Pictures," *J. Royal Stat. Soc. Ser.B,* Vol. 48, 1986, pp. 259-302.
3. Chellappa, R., "Two-Dimensional Discrete Gaussian Markov Random Field Models for Image Processing," *Progress in Pattern Recognition 2* edited by L.N. Kanal and A. Rosenfeld, North-Holland Publ. Co., 1985, pp. 79-112.
4. Chen, C.C., *Markov Random Fields in Image Analysis,* Ph.D. Thesis, Michigan State University, East Lansing, 1988.
5. Chow, P.B., Brown, C.M., and Raman, R., "A Confidence-Based Approach to the Labeling Problem," *IEEE Workshop on Image Understanding,* Miami, 1987, pp. 51-56.
6. Cohen, F.S. and Cooper D.B., "Simple Parallel Hierarchical and Relaxation Algorithms for Segmenting Noncausal Markovian Random Fields," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 9, 1987, pp. 195-219.

7. Derin, H. and Elliott, H., "Modeling and Segmentation of Noisy and Textured Images Using Gibbs Random Fields," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 9, 1987, pp. 39-55.
8. Dubes, R.C., Jain, A.K., Sateesha, S.G., and Chen, C.C., "MRF Model-Based Algorithms for Image Segmentation," *in Proc. IEEE International Conference on Pattern Recognition,* 1990, pp. 808-814.
9. Geman, S. and Geman, D., "Stochastic Relaxation: Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 6, 1984, pp. 721-741.
10. Haralick, R.M. and Shapiro, L.G., "Survey: Image Segmentation," *Computer Vision, Graphics, and Image Processing,* Vol. 29, 1985, pp. 100-113.
11. Kindermann, R. and Snell, J.L., *Markov Random Fields and Their Applications,* American Mathematical Society, Volume I, 1980.
12. Mardia, K.V. and Hainsworth, T.J., "A Spatial Thresholding Method for Image Segmentation," *IEEE Trans. Pattern Anal. Machine Intell.,* Vol. 10, 1988, pp. 919-927.
13. Marroquin, J., Mitter, S., and Poggio, T., "Probabilistic Solution of Ill-Posed Problems in Computational Vision," *J. American Statistical Association,* Vol. 82, 1987, pp. 76-89.
14. Pappas, T.N. and Jayant, N.S., "An Adaptive Clustering Algorithm for Image Segmentation," *in Proc. IEEE International Conference on Computer Vision,* 1988, pp. 310-315.
15. Rosenfeld A. and Kak, A.C., *Digital Picture Processing,* Vol. 2, New York: Academic Press, 1982.

APPENDIX A
MRF and the posteriori distribution

Let $P(\mathbf{x}) = f(\mathbf{x}|\mathbf{y}) = f(\mathbf{x}|\mathbf{y})/f(\mathbf{y})$, $x_i \in A$ for each i, and assume

(1) $f(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{MN} f(y_i|x_i)$,

(2) $f(x_t|x_r, r \neq t) = f(x_t|\mathbf{x}_{\partial t})$, namely, $\mathbf{x}$ is a realization of an MRF.

Then, "P" defines an MRF for $\mathbf{X}|\mathbf{y}$.

Proof: Want to prove $P(x_s|\tilde{\mathbf{x}}_s) = P(x_s|\mathbf{x}_{\partial s})$.

$$P(x_s|\tilde{\mathbf{x}}_s) = \frac{P(x_s, \tilde{\mathbf{x}}_s)}{\sum_{g \in A} P(x_s = g, \tilde{\mathbf{x}}_s)} = \frac{f(x_s, \tilde{\mathbf{x}}_s|\mathbf{y})}{\sum_{g \in A} f(x_s = g, \tilde{\mathbf{x}}_s|\mathbf{y})}$$

$$= \frac{f(\mathbf{y}|x_s, \tilde{\mathbf{x}}_s)\ f(x_s|\tilde{\mathbf{x}}_s)}{\sum_{g\in A}\left[f(\mathbf{y}|x_s=g, \tilde{\mathbf{x}}_s)\ f(x_s=g|\tilde{\mathbf{x}}_s)\right]} = \frac{\{\prod_{i=1,i\neq s}^{MN} f(y_i|x_i)\}\ f(y_s|x_s)\ f(x_s|\tilde{\mathbf{x}}_s)}{\sum_{g\in A}\left[\{\prod_{i=1,i\neq s}^{MN} f(y_i|x_i)\}\ f(y_s|x_s=g)\ f(x_s=g|\tilde{\mathbf{x}}_s)\right]}$$

$$= \frac{f(y_s|x_s)\ f(x_s|\tilde{\mathbf{x}}_s)}{\sum_{g\in A}\left[f(y_s|x_s)\ f(x_s=g|\tilde{\mathbf{x}}_s)\right]} = \frac{f(y_s|x_s)\ f(y_s|\mathbf{x}_{\partial s})}{\sum_{g\in A}\left[f(y_s|x_s=g)\ f(x_s=g|\mathbf{x}_{\partial s})\right]}$$

On the other hand, let $T = S-\{s\}-\{\partial s\}$ and denote all x_i, $i\in T$ by x_T; then

$$P(x_s|\mathbf{x}_{\partial s}) = \frac{\sum_{x_T} P(\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)}{\sum_{x_s\in A}\left[\sum_{x_T} P(\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)\right]} = \frac{\sum_{x_T} f(\mathbf{y}|\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)\ f(\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)}{\sum_{x_s\in A}\left[\sum_{x_T} f(\mathbf{y}|\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)\ f(\mathbf{x}_{\partial s}, x_s, \mathbf{x}_T)\right]}$$

$$\text{Denominator} = \left[\sum_{x_s\in A}\ \sum_{x_T}\{[\prod_{i\in T} f(y_i|x_i)]\ [\prod_{\tau\in\partial s} f(y_\tau|x_\tau)]\ f(y_s|x_s)\ f(x_s|\tilde{\mathbf{x}}_s)\ f(\tilde{\mathbf{x}}_s)\}\right]$$

$$= \left[\prod_{\tau\in\partial s} f(y_\tau|x_\tau)\right]\left[\sum_{x_\tau}\{[\prod_{i\in T} f(y_i|x_i)f(\tilde{\mathbf{x}}_s)\}\right]\left[\sum_{x_s\in A} f(y_s|x_s)f(x_s|\mathbf{x}_{\partial s})\right]$$

$$\text{Numerator} = \sum_{x_T}\left[\prod_{i\in T} f(y_i|x_i)]\ [\prod_{\tau\in\partial s} f(y_\tau|x_\tau)]\ f(y_s|x_s)\ f(x_s|\tilde{\mathbf{x}}_s)\ f(\tilde{\mathbf{x}}_s)\right]$$

$$= \left[\prod_{\tau\in\partial s} f(y_\tau|x_\tau)\right]\left[\sum_{x_\tau}\{\prod_{i\in T} f(y_i|x_i)f(\tilde{\mathbf{x}}_s)\}\quad f(y_s|x_s)f(x_s|\mathbf{x}_{\partial s})\right]$$

Then,

$$P(x_s|\mathbf{x}_{\partial s}) = \frac{f(y_s|x_s)f(x_s|\mathbf{x}_{\partial s})}{\sum_{x_s\in A}[f(y_s|x_s)\ f(x_s|\mathbf{x}_{\partial s})]} \tag{1}$$

Therefore, $P(x_s|\tilde{\mathbf{x}}_s) = P(x_s|\mathbf{x}_{\partial s})$, namely, "P" defines an MRF.

APPENDIX B
GMRF Sampling Algorithm

A Gaussian Markov random field (GMRF), defined on a lattice L, can be viewed as a multivariate normal distribution [Bes74] with a mean vector $\mu\mathbf{1}$ consisting of all μ's and a covariance matrix σ^2B [1], where the correlation matrix B is block-circulant. For a 2nd-order GMRF, B was given in *Case 2*, Section 2. There are six parameters, μ, σ, θ_1, θ_2, θ_3, and θ_4 in a 2nd-order GMRF. A sampling algorithm, proposed by Chellappa [3], is given below.

/* Generate an M×N image from a GMRF; A(i,j) = B(1, j + (i − 1)N)*/
(a) Generate an M×N array η with each element i.i.d. from $N(0,\sigma^2)$,
(b) Apply 2-D Fourier transform on η, save the result in η,
(c) Apply 2-D inverse Fourier transform on A, save the result in A,
(d) $x(u,v) \leftarrow \eta(u,v)/\sqrt{A(u,v)}$, for u = 0,1,....,M-1; v = 0,1,....,N-1.
(e) Apply 2-D inverse Fourier transform on **x**, save the result in **x**.
(f) $\mathbf{x} + \mu\mathbf{1}$ is a realization.

Chaur-Chin Chen (陳朝欽) was born in Kaohsiung, Taiwan, in 1955. He received a B.S. degree in mathematics from National Taiwan University, Taipei, Taiwan, in 1977, M.S. degrees in mathematics and computer science in 1982 and 1984, and a Ph.D. degree in computer science in 1988 from Michigan State University.

He is currently an Associate Professor in the Department of Computer Science at National Tsing Hua University, Taiwan, R.O.C. His research interests are in the areas of image pattern recognition, computer vision, and image models.

Dr. Chen is a member of the Pattern Recognition Society, the Association for Computing Machinery, and IEEE Computer Society.

Richard C. Dubes was born in Chicago, IL. He received a B.S. degree from the University of Illinois in 1956 and M.S. and Ph.D. degrees from Michigan State University, East Lansing in 1959 and 1962, respectively, all in electrical engineering.

He is currently a Professor in the Department of Computer Science at Michigan State University.

Dr. Dubes is a member of the Pattern Recognition Society, the Classification Society and Sigma Xi, and is an Associate Editor of Pattern Recognition.

Volume 6, Number 4, December 1990
JOURNAL OF INFORMATION SCIENCE AND ENGINEERING

CONTENTS

PAPERS